Award Number: DAMD17-02-1-0499

TITLE: Cloning and Characterization of Expanded CAG-Repeat Containing Sequence(s): Identification of Candidate Breast Cancer Gene(s)

PRINCIPAL INVESTIGATOR: Hilmi Ozcelik, Ph.D.

CONTRACTING ORGANIZATION: Mount Sinai Hospital
Toronto, Ontario, Canada M5G 1X5

REPORT DATE: June 2003

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2003	Annual (1 May 02 - 30 May 03)

4. TITLE AND SUBTITLE
Cloning and Characterization of Expanded CAG-Repeat Containing Sequence(s): Identification of Candidate

5. FUNDING NUMBERS
DAMD17-02-1-0499

6. AUTHOR(S)
Hilmi Ozcelik, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Mount Sinai Hospital
Toronto, Ontario, Canada M5G 1X5

E-Mail: ozcelik@mshri.on.ca

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates: All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

Genetic factors have been shown to influence the risk of developing breast cancer. To date only two breast cancer predisposition genes, BRCA1 and BRCA2, have been identified. A wide spectrum of BRCA mutations has been found in breast cancer families. However, in the majority of breast cancer cases BRCA1 or BRCA2 mutations were not detected, suggesting the presence of other breast cancer predisposition genes. Trinucleotide repeats are present widely in human genome. Expansions of some trinucleotide repeats have been recognized to be the cause of several genetic disorders. In a previous study we have utilized the Repeat Expanded Detection (RED) method to investigate the distribution of (CAG) repeats in a sample of 212 breast cancer and 196 normal population controls and in 20 breast cancer mother-daughter pairs. We have shown that 2.4% of the breast cancer cases revealed repeat sizes of over (CAG)144. No expansion of this magnitude has been detected in 196 population controls samples. Among the mother–daughter pairs one (5.0%) has shown a CAG expansion that was detected in daughter but not in mother older generation. RED analysis does not provide information on the location of expansions in the genome, therefore it is necessary to clone and localize the expanded repeat regions. At this stage of the project, we have developed an efficient cloning strategy, and validated the results by sequencing the inserts. We have shown that the second enrichment step has drastically improved the detection of fragments with CAG repeats. We have also detected a large GAG/CAG containing fragment of 100 repeats suggesting that our system is able to identify large repeat containing inserts. Currently, we are at the stage of further validation and application of the strategy to breast cancer specimens.

14. SUBJECT TERMS
Breast cancer, CAG/CTG repeats, Rapid Expansion Detection, Trinucleotide Repeat Expansions, Cloning.

15. NUMBER OF PAGES
13

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION:

Expression of an inherited disease gene does not always follow a Mendelian inheritance pattern. Since the turn of the century, it has been observed that the transmission of certain disease genes from one generation to the next is associated with an increase in the severity of the disease symptoms and/or a decrease in the age of onset. This clinical phenomenon is called anticipation and it was observed with many inherited diseases, especially ones involving neurological disorders. However, the absence of a satisfactory physical explanation for anticipation at the molecular level resulted in much controversy in the past over the acceptance of the concept of genetic anticipation.

Advances from molecular genetic analysis of many of the inherited human diseases in which anticipation has been observed has recently attributed anticipation in these diseases to a new class of dynamic mutations, characterized by trinucleotide repeat expansions in the locus where the disease genes have been mapped. The effect depends on the location of the repeats relative to the gene and the type of repeats (for review see Sanjeeva et. al., 1997, Margolis et al., 1999, Vincent et al., 2000). The molecular consequences that result from trinucleotide repeat expansions and the mechanism by which they lead to pathology may be quite diverse. In general however, trinucleotide repeat expansions have been shown to perturb either the structure and function (type I mutations) or the expression (type II mutations) of the affected gene (for review see Sanjeeva et. al., 1997, Margolis et al., 1999, Vincent et al., 2000). For example, expansion of a (CAG)n-repeat in the coding region of the Huntington's disease gene allows expression of an altered protein which contains an expanded polyglutamine region, resulting in altered conformation, processing and general physical properties of the protein function of the product (Trottier et al., 1995). Alternatively, expansion which results in 200-2000 CGG repeats in the 5' untranslated region of the Fragile X syndrome (FRAXA) gene result in the loss of expression of FRAXA mRNA (Pieretti et. al.,1991). As the repeat expands with transmission to the next generation, the CGG repeats become more methylated reducing transcription of the FRAXA gene. The nature of this dynamic group of mutations, which can involve very large amplification of trinucleotide repeats, renders the sequence unstable during meiosis. This results in intergenerational instability of the length of the trinucleotide repeat. It is not known why repeats which exceed a critical value are unstably transmitted to succeeding generations with a tendency towards expansion of trinucleotide repeats. There is however, a very clear association between longer expansions at the disease locus in the succeeding generations and earlier clinical manifestation (Sanjeeva et. al., 1997, Margolis et al., 1999, Vincent et al., 2000). Anticipation, therefore, is now commonly accepted as a hallmark of the inheritance of an amplified trinucleotide repeat expansion mutation. Thus far, at least 12 genetic diseases (mostly muscular or neurological disorders) have been attributed to expansions of trinucleotide repeats in the loci containing the disease gene. These diseases are characterized as having increasing copy numbers of the unstable expanded sequences with subsequent generations.

One can envision that expansion of trinucleotide repeats are not restricted to neuro-muscular disorders, and they probably represent a novel class of dynamic mutations causing various human diseases. A study by O'Donovan et al., 1996,

provided an intriguing finding suggesting that repeat expansions have a wide spread role in common human diseases. This study has shown that older, healthy individuals have generally shorter CAG-repeat lengths than their younger counterparts. The demonstration of decreased, genome wide repeat-copy number with age in healthy populations suggests that the dynamic mutation could have a wide-spread role in human susceptibility to disease. These findings also suggest that trinucleotide repeat diseases do not display a single major gene inheritance.

Statement of Work

Technical Objective 1: Cloning of Gene Sequences with CAG-Repeat Expansion.
(Four parallel cloning procedures will be performed simultaneously using 4 pools each including DNA mixed from 2 RED positive cases)

Task 1: Months 1-4	Enrichment by digestion of DNA samples and separation on agarose gels. Slicing (2-4mm), and extracting the DNA from each slice (approximately 200-400 slices). RED analysis of all 200-400 slices to locate the fragments with expansion.
Task 2: Months 4-6	Cloning into ZapII system, electroporation into E.Coli, and amplification. Secondary enrichment by pooling: Plating the bacteria, pooling and amplification of pools. Extraction of DNA (~100-200) and RED analysis (~ 100-200).
Task 3: Months 6-8	Transformation of the RED positive DNA into E.Coli CJ236 strain and generation of ssDNA. Isolation of the ssDNA and production of dsDNA containing only CAG inserts CAG-probes and the primer extension method. Electroporation into E.Coli, amplification, and plating. Selection and amplification of individual clones. Extraction of DNA from individual clones (~100-200).
Task 4: Months 8-10	RED analysis on the DNA extracted from the individual clones (~100-200). Sequencing of inserts from RED positive clones. Identification of clones with large repeat sequences.
Task 5: Months 10-11	Designing PCR primers for every sequence with large CAG-repeats identified through cloning. Screening of all the cases originally detected to have the CAG-repeat expansion (detected by RED) by PCR and sequencing.

Technical Objective 2:	**Identification and Characterization of Gene(s) containing or Flanking Expanded CAG-Repeats**
Task 1: Months 11-12	Designing PCR protocols for all the sequences identified to have large CAG-repeats through cloning and optimization of their detection through microsatellite analysis. Sequencing of a small panel of control specimens to identify repeats with

varying sizes to be used as size controls on microsatellite gels.

Task 2: Months 12-17 — Microsatellite analysis of 100 control specimens for every repeat cloned to determine the allelic frequency of each repeat. It is expected than several regions containing large CAG-repeats will be identified from cloning of RED positive DNA from 8 breast cancer cases.

Task 3: Months 17-22 — Extraction of sequence data using different sequence database sources including GenBank, EST and STS databases and putting together the pieces of information to find longer sequence information. Mapping of the CAG repeats using public genetic map information, and identify other genes located around the CAG-repeat. Locating the exact position of CAG repeat in relation to the structure of the genes identified.

Task 4: Months 20-24 — Obtaining the sequence data from the databases and interpretation of this information with the presence of an extensive literature search. If necessary, wet lab work will be carried out to clarify the position of CAG-repeat in relation to the structure of gene(s) identified. Search for information published on the loci/genes found and their association with cancer specifically with breast cancer. Writing of manuscripts to be proposed to peer reviewed journals.

Body

1. Overview

Our goal is to clone and characterize expanded CAG repeat containing sequences in breast cancer patients. These genes may represent breast cancer predisposition genes. Most of the reported methods utilized for library contraction and cloning of the large repeat containing fragments require the use of large amount of genomic DNA as a starting template (Yuan et al, 2001; Vincent et al, 2000; Koob et al, 1998). Since patient material is a limited source, we were specifically interested in developing an efficient cloning strategy that enriches for long CAG repeat containing fragment with the minimum use of starting genomic DNA. Long trinucleotide repeats tends to form special secondary structures, which reduce the efficiency of cloning of DNA fragments containing expanded repeats. Therefore, cloning of long repeat containing fragments is cumbersome. (Koob et al, 1998, Sanpei et al, 1996)

2. Establishment of the Cloning Procedure for Repeat Containing Sequences

2A. Enrichment for Long CAG-Repeat Containing Fragments

Large CAG repeat containing genomic DNA sample was digested with Sau3AI. Adaptors complementary to Sau3AI cut sites were ligated to digested product. The ligated product was used as template for a PCR reaction using one of the adaptors as a primer. The PCR reaction resulted in a smear of amplified fragments that ranged in length from 100 bp to more than 3 kb **(Figure 1).** Several PCR conditions have been applied and subjected to the further stages of the procedure.

The PCR product was purified and a biotin labeled $(CAG)_8$ oligo was annealed to the product and a nucleotide biased extension step was performed using dATP, dCTP and dGTP. The extension reaction product was mixed with Streptavidin MagneSphere Paramagnetic Particles and then single strand DNA fragments were eluted and purified by ethanol precipitation. PCR was performed using the eluted fragments as template and the same adaptor as primer. The PCR product was then digested with Sau3AI and new adaptors were linked to the digested product and the whole enrichment process was repeated **(Figure 2).** The final PCR product was cloned into a T vector using the Pgem T easy vector system II kit (Promega) and the XL1-Blue MRF' bacterial strain.

This strain was chosen because of its tolerance to host vectors containing fragments with long repetitive DNA. White colonies were picked and cultured for 6 hours in 96 well plates containing 200ul of LB-Ampicillin medium. PCR was performed using the cultured bacteria directly as template and theT7/SP6 primer set **(Figure 3).** We noticed that most of the PCR products were very short which means that the cloning process was in favor of short fragments.

2B. Enrichment by Gel Fragmentation

Since the presence of small PCR products in the sample resulted in competing out the cloning of larger fragments. We have developed the gel fragmentation system before the cloning step where shorter fragments (<300bp) were removed. The final PCR product was ran on a 1% agarose gel and the gel was divided into 3 pieces: short

fragments (<300 bp), medium fragments (300-2000 bp), and long fragments > 3000 bp **(Figure 4)**.

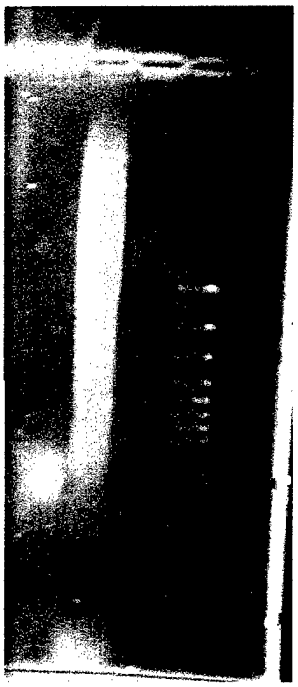

Figure 1. PCR product of digested DNA and ligated to adaptors

The DNA was recovered from the short, medium and long portions of the gel and cloned into the T vector. Cloned fragments were amplified by using the vector specific primers (SP6/T7) **(Figure 5)**.

2C. PCR-Based Colony Screening

Initially, we have sequenced multiple colonies to assess the type of fragments incorporated into the vectors. In order to increase the efficiency of screening inserts we have developed a PCR-based screening method. The individual or pooled colonies can be lyzed and directly amplified. This is an efficient method to screen for colonies that contain large repeats, eliminating the colonies containing short inserts. This method will be used to enrich for colonies having larger (potentially the large repeat) inserts, which will be then subjected to direct sequencing for validation.

2D. Enhancement of Percentage of CAG Repeat Containing Fragments

DNA sequencing was performed on 24 PCR products resulting from cloning of fragments after one round of enrichment and on 32 PCR product resulting from cloning of fragments after two round of enrichment. As shown in **Table1** repeating the enrichment process for another round prior to cloning has enhanced the percentage of CAG containing fragments from 17% to 66 %. Most of he detected CAG repeats

ranged in size between $(CAG)_4$ and $(CAG)_{12}$. We were able to detect two repeats larger than $(CAG)_{12}$; one with $(CAG)_{13}$ and one with $(CAG)_{17}$.

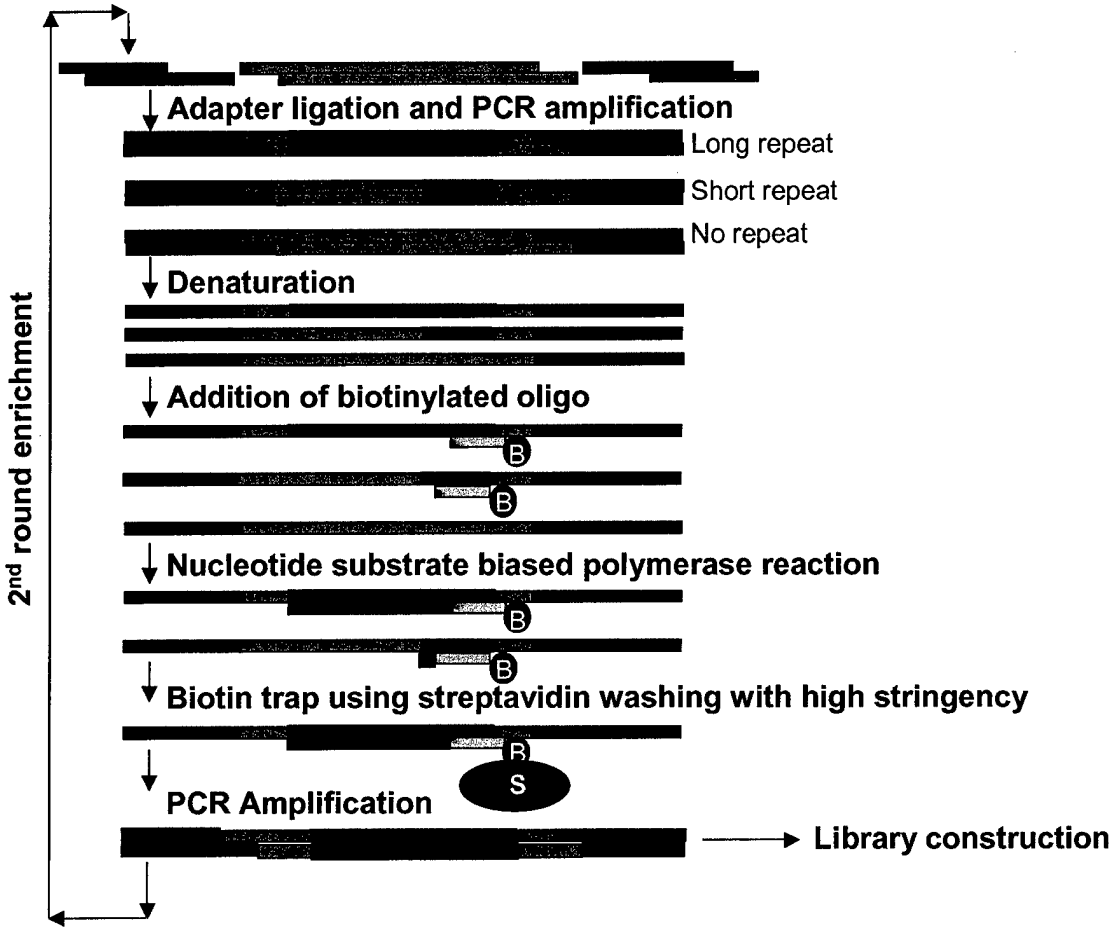

Figure 2: Schematic representation of enrichment for long CAG repeat containing fragments by using magnetic bids.

2E. Library Construction and Screening

We are now in the process of library constructing and screening using the above mentioned enrichment and cloning strategy. Using this approach we will be able to identify the location and the flanking sequences of the large CAG repeats which were detected by RED in our previous project.

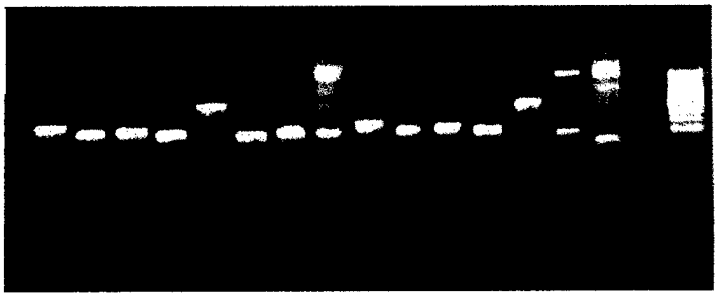

Figure 3. PCR products of amplified plasmids using t7/SP6 primers

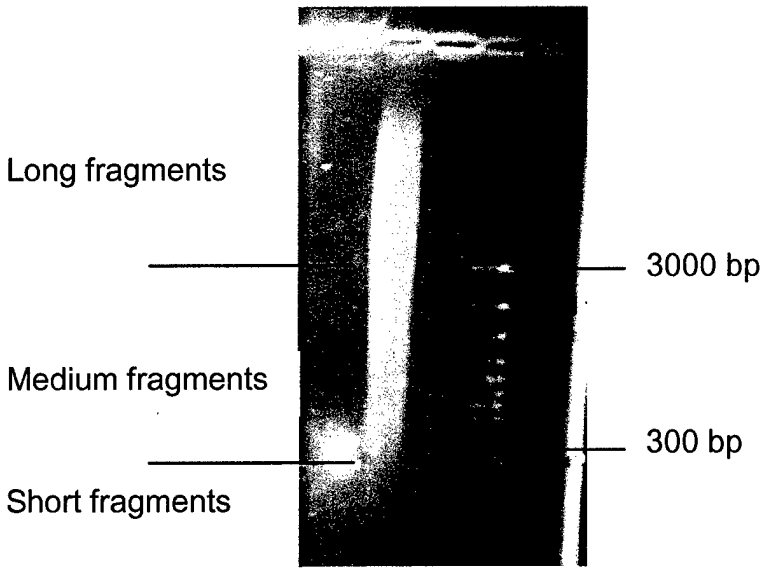

Figure 4. Enrichment by gel fragmentation

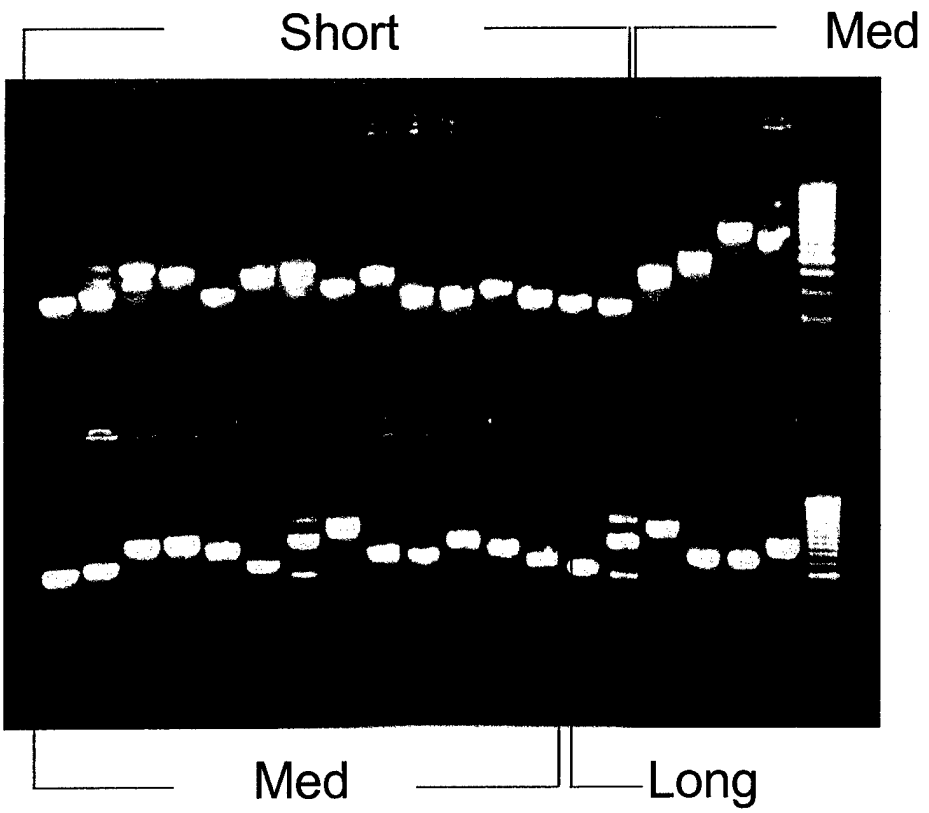

Figure 5. PCR products of amplified plasmids using t7/SP6 primers after gel fragmentationfragmen

	(CAG)n containing clones	no (CAG)n containing clones	Total	% of (CAG)n containing clones
One round enrichment	4	20	24	17
Two rounds enrichment	21	11	32	66

Table 1. Enhancement of percentage of CAG containing clones by enrichment for repeats containing fragments

3. Immediate Future Task

Using the above optimized cloning strategy we will complete the screening for RED positive clones in order to identify expanded repeat and their flanking sequences. In case we are not able to detect any long repeat containing clones the system needs to be retested for its efficiency for cloning long CAG containing fragments. If necessary we will seek collaboration with well established laboratory to complete the cloning process.

4. Key Research Accomplishments

At this stage, we have developed a cloning strategy by integrating several enrichment steps for efficient identification of large repeat containing fragments. The specific technical accomplishments include:

(a) Enrichment for CAG repeat containing fragments using magnetic beads
(b) Enrichment by gel fragmentation
(c) Enhancement of the percentage of CAG repeat containing clones
(d) Development of a PCR based colony-screening method

We have also carried out a validation study to assess the efficiency of our strategy to detect large repeat containing fragments.

5. Reportable Outcomes

Hamdi Jarjanazi, Hong Li, Irene L. Andrulis, and Hilmi Ozcelik, "Investigating cag/ctg trinucleotide repeat lengths in breast cancer", Manuscript under revision by the Journal of Medical genetics.

Hamdi Jarjanazi, Hong Li, Irene L. Andrulis, and Hilmi Ozcelik, "Cloning strategy for trinucleotide repeat expansion: Searching for novel breast cancer predisposition gene(s)". Controversies in the etiology, detection and treatment of breast cancer : 2002, June 13-14, 2002 Toronto, Ontario, Canada *(Poster)*.

Hamdi Jarjanazi, Hong Li, Irene L. Andrulis, and Hilmi Ozcelik, "Identification of novel breast cancer predisposition gene(s) with trinucleotide repeat expansions". The Samuel Lunenfeld Research Institute Annual Retreat: 2002: October, 9-10 YMCA Geneva Park, Ontario, Canada *(Poster)*.

Hamdi Jarjanazi, Hong Li, Irene L. Andrulis, and Hilmi Ozcelik, "Identification of novel breast cancer predisposition gene(s) with trinucleotide repeat expansions". AACR 2003 Annual Meeting: 2002, April 5-9, Toronto, Ontario Canada *(Mini-symposium)*.

References

Koob MD, Benzow KA, Bird TD, Day JW, Moseley ML, Ranum LP. Rapid cloning of expanded trinucleotide repeat sequences from genomic DNA. Nat Genet 1998 Jan;18(1):72-5

Margolis RL, McInnis MG, Rosenblatt A, Ross CA. Trinucleotide repeat expansion and neuropsychiatric disease. Arch Gen Psychiatry 56(11):1019-31, 1999.

O'Donovan MC, Guy C, Craddock N, Bowen T, McKeon P, Macedo A, Maier W, Wildenauer D, Aschauer HN, Sorbi S, Feldman E, Mynett-Johnson L, Claffey E, Nacmias B, Valente J, Dourado A, Grassi E, Lenzinger E, Heiden AM, Moorhead S, Harrison D, Williams J, McGuffin P, Owen MJ.Confirmation of association between expanded CAG/CTG repeats and both schizophrenia and bipolar disorder. Psychol Med 26(6):1145-53, 1996.

Pieretti M, Zhang FP, Fu YH, Warren ST, Oostra BA, Caskey CT and Nelson DL. Absence of expression of the FMR-1 gene in fragile X syndrome. Cell 66:817-822, 1991.

Sanjeeva P and Housman DE. The complex pathology of trinucleotide repeats. Current Opinion in Cell Biology 9:364-372, 1997.

Sanpei K, Takano H, Igarashi S, Sato T, Oyake M, Sasaki H, Wakisaka A, Tashiro K, Ishida Y, Ikeuchi T, Koide R, Saito M, Sato A, Tanaka T, Hanyu S, Takiyama Y, Nishizawa M, Shimizu N, Nomura Y, Segawa M, Iwabuchi K, Eguchi I, Tanaka H, Takahashi H, Tsuji S. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. Nat Genet. 1996 Nov;14(3):277-84.

Trottier Y, Lutz Y, Stevanin G, Imbert G, Devys D, Cancel G, Saudou F, Weber C, David G, Tora L, et al. Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. Nature 378:403-406, 1995.

Vincent JB, Neves-Pereira ML, Paterson AD, Yamamoto E, Parikh SV, Macciardi F, Gurling HM, Potkin SG, Pato CN, Macedo A, Kovacs M, Davies M, Lieberman JA, Meltzer HY, Petronis A, Kennedy JL., An unstable trinucleotide-repeat region on chromosome 13 implicated in spinocerebellar ataxia: a common expansion locus Am J Hum Genet 2000 Mar;66(3):819-29

Yuan QP, Lindblad-Toh K, Zander C, Burgess C, Durr A, Schalling M. A, cloning strategy for identification of genes containing trinucleotide repeat expansions. Int J Mol Med. 2001 Oct;8(4):427-31